Lecture 10

Gauss Seidel via matrices
Then Gauss Seidel vs. Coordinate Descent

Onto Eigenvalues
and the power method!

Where to eigenvalues occur?
A cute example from phase retrieval

The power method converges
proportional to the difference in
top eigenvalues, which follows
from the spectral decomposition
of a matrix.

- Twelve input = twelve
  9. (A)
  ↳ Check notes online
  ↳ Should be online soon.

Gauss Seidel is a
by - Tuesday!

$$x_1^{(k+1)} = \frac{1}{A_{11}}\left(b_1 - \sum_{j>1} A_{1j} x_j^{(k)}\right)$$

$$x_2^{(k+1)} = \frac{1}{A_{22}}\left(b_2 - \sum_{j\neq 2} A_{2j} x_j^{(k)} - \sum_{j>2} A_{2j} x_j^{(k)}\right)$$

$$x_i^{(k+1)} = \frac{1}{A_{ii}}\left(b_i - \sum_{j<i} A_{ij} x_j^{(k+1)} - \sum_{j>i} A_{ij} x_j^{(k)}\right)$$

$$x^{(k+1)} = D_n(b_n - U_n x^{(k)}) - L_n x^{(k+1)}$$

so by

$$(L_n)_{ij} = \begin{cases} A_{ij}, & j<i \\ 0 & \text{else.} \end{cases}$$

$$(L_n)_{ij} = \begin{cases} A_{ij}, & j<i \\ 0 & \text{else.} \end{cases}$$

$$x^{(k+1)} = D_n(b_n - U_n x^{(k)})$$

$$\Leftrightarrow$$

$$x^{(k+1)} + D_n L_n x^{(k+1)} = D_n(b_n - U_n x^{(k)})$$

$$(I + D_n L_n)x^{(k+1)} = D_n(b_n - U_n x^{(k)})$$

$$x^{(k+1)} = (I + D_n L_n)^{-1}(D_n(b_n - U_n x^{(k)}))$$

$$= \left[(D_n(D_n^{-1} + L_n))^{-1}(D_n)\right](b_n - U_n x^{(k)})$$

$$= (D_n^{-1} + L_n)^{-1}(b_n - U_n x^{(k)})$$

Using the same
analogy - you can
$\rho(\,(D_n^{-1} + L_n)^{-1}(U_n)\,) < 1$
then the converge.

Gauss Seidel $\Leftrightarrow$
Coordinate Descent.

- The hard! under
  special condition.

Can I prove $\Rightarrow$

$$f(x) = \frac{1}{2}x^T A x - x^T \frac{b}{2}$$

$$x_{k+1} = \frac{\xi_k + \gamma g_k e_i}{\sqrt{\|A e_i\|_2^2}}$$

Gauss Seidel ($\gamma = 1$)

$$x_i^{(k+1)} = \frac{1}{A_{ii}}\left(b_i - \sum_{j>1} A_{ij} x_j^{(k)}\right)$$

$$x_i^{(k+1)} = \frac{1}{A_{ii}}\left(b_i - \sum_{j\neq i} A_{ij} x_j^{(k)}\right)$$

Suppose A has
1's on the diag

$A_{ii} = 1$

$$\sum_{j\neq i} A_{ij} x_j^{(k)} + x_i^{(k)} = b_i$$

$$x_i^{(k)} = x_i^{(k)} - \sum_{j\neq i} A_{ij} x_j^{(k)} + b_i$$

$$= b_i - \sum_{j\neq i} A_{ij} x_j^{(k)}$$

GS $x^{(k+1)} = b_i - \sum_{j\neq i} A_{ij} x_j^{(k)}$ $\Rightarrow$ $x_i^{(k+1)} = b_i - \sum_{j\neq i} A_{ij} x_j^{(k)}$

(Conv. $\Leftrightarrow$ $\rho(\cdot) < 1$)

- Eigenvalues are
  unique up to
  scaling.

$$Ax = \lambda x \text{ then}$$

$$A(\alpha x) = \alpha(Ax)$$

$$= \alpha\lambda x$$

$$= \lambda(\alpha x).$$

$\Rightarrow$ If $A = A^T$, there
are unique U, Λ

$$U = [u_1, \dots, u_n],$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$AU = U\Lambda$$

$$\Leftrightarrow$$

$$A = U\Lambda U^T$$

Also $UU^T = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$
↳ Orthogon.

$$A U = U\Lambda \Leftrightarrow$$

$$\boxed{A = U\Lambda U^T.}$$

Called spectral
decomp.

- $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is
  non-zero, but
  $\lambda_1 = 0$.

- A sym. matrix
  always has an
  eigenbasis. (Real)
  and an orthogon
  Eigenbasis.

- Eigenvalues are
  not unique.
  $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is
  non-zero, but
  $\lambda_1 = 0$.

- A sym. matrix
  always has an
  eigenbasis. (Real)
  and an orthogon
  Eigenbasis.

- Eigenvalues are
  not unique.
  $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

How to get an Eigenpair:
Have $\frac{x^T A x}{x^T x}$ s.t. $\|x\|_2 = 1$

$$f(x, \lambda) = x^T A x - \lambda(x^T x - 1)$$

$$\frac{\partial f}{\partial \lambda} = \partial A x - \partial \lambda x = 0 \Rightarrow Ax = \lambda x$$

$$\frac{\partial f}{\partial x} = x^T x - 1 = 0 \Rightarrow x^T x = 1$$

$$\Leftrightarrow$$

Stationary points are Eigenpairs/values
pairs

Idea for the
Use Gradient Ascent on
$\frac{x^T A x}{x^T x}$ s.t. $\|x\|_2^2 = 1$.

Gradient Ascent

$$x_{k+1} = \xi_k + \gamma \underbrace{\nabla_x f(x_k)}_{\text{grad}}$$

step. grad

Problem

So we find the closest vector
to $x_k + \gamma g \partial A x_k$
with $\|\cdot\|_2^2 = 1$.

This is given by

$$\frac{x + \gamma g \partial A x_k}{\|x + \gamma g \partial A x_k\|_2}$$

$\Rightarrow$ Projected
gradient ascent.

One rule: if
$\gamma \to \infty$.
then we get

$$\boxed{\frac{A x_k}{\|A x_k\|}}$$

called the Power
method.

The power method for
Eigenpairs/eigenvalues

x_0 = random

$$x_{k+1} = \frac{A x_k}{\|A x_k\|_2}$$

$$\lambda_{k+1} = x_k^T A x_k$$

Claim: A^k Let $\|x\|_2 = 1$

Lemma: (clear.)

$$x_k = \frac{A x_0}{\|A x_0\|_2}$$

$$= \frac{A^k x_0}{\|A^k x_0\|}$$

$$= \rho_k \cdot A x_0.$$

$$x_k = \frac{A x_1}{\|A x_1\|_2} = \rho_k A^k x_0$$

$$= \rho_{k+1}$$

Induction by:

$$x_{k+1} = \frac{A x_k}{\|A x_k\|_2}$$

$$= \frac{A \rho_k A^k x_0}{\|A \rho_k A^k x_0\|}$$

$$= \frac{\boxed{\rho_k}}{\|\rho_k A^{k+1} x_0\|} \cdot A^{k+1} x_0$$

$$= \rho_{k+1}$$

$$A^k = U \Lambda^k U^T$$

So

CS 510 - Purdue University - Computer Science - David F. Gleich - 2019